

THE UTILIZATION OF INDIVIDUAL CAPABILITIES IN GROUP APPROACHES TO STRATEGIC DECISION-MAKING

DAVID M. SCHWEIGER and WILLIAM R. SANDBERG

College of Business Administration, University of South Carolina, Columbia, South Carolina, U.S.A.

Previous research has found that groups using dialectical inquiry (DI) or devil's advocacy (DA) make better strategic decisions than groups using a consensus (C) approach. This paper explains those findings by using new data to show that the DI and DA approaches make better use of the capabilities of individual group members. Specifically, the DI and DA groups yielded significantly higher quality recommendations and assumptions than the average of the individuals in the respective groups, whereas the C groups did not. Moreover, the recommendations and assumptions of the DI groups and the recommendations of the DA groups significantly exceeded those of the best individual in the respective groups. There were no significant differences for the C groups.

During the past decade theorists and researchers have debated and studied two approaches for structuring strategic decision-making groups: dialectical inquiry (DI) and devil's advocacy (DA) (Cosier, 1978, 1980; Cosier and Aplin, 1980; Cosier, Ruble and Aplin, 1978; Mitroff, Barabba and Kilmann, 1977; Schwenk 1982b, 1984; Schwenk and Cosier, 1980; Cosier and Rechner, 1985; Schweiger, Sandberg and Ragan, 1986). DI and DA build structured, constructive *interpersonal* conflict into the decision processes of top management groups facing complex, ill-defined (i.e. wicked) strategic problems. Both approaches aim to improve decisions by requiring group members to critically examine each other's and the group's recommendations and assumptions on which the recommendations depend (Mason and Mitroff, 1981).

Although both DI and DA rely on interpersonal conflict as their primary structural mechanism, there are differences between them that suggest possible differences in their effectiveness. DI uses a structured debate among two sets of group members who represent diametrically opposed recommendations and assumptions, whereas DA uses a structured critique by one set of group members of recommendations and assumptions

developed by a second set as bases for critical examination (Mitroff, 1982b; Mason and Mitroff, 1981; Mitroff, 1981; Mitroff and Mason, 1981). According to Mason (1969), DI should lead to higher-quality solutions than DA because it seeks to identify alternatives from the original set of diametric recommendations and assumptions, whereas DA focuses only on what is wrong with recommendations and assumptions, rather than on identifying suitable alternatives.

A number of studies have attempted to assess the comparative effectiveness of dialectical inquiry and devil's advocacy. On the basis of two field studies (Lourenco and Glidewell, 1974; Mitroff *et al.*, 1977) Mitroff and his associates have argued for the superiority of DI (Mitroff and Mason, 1981; Mitroff, 1982a,b). Criticizing the two field studies for their inability to establish causality, Cosier and his associates have conducted a series of laboratory studies that questioned the superiority of DI (Cosier, 1981, 1982, 1983; Schwenk, 1982a, 1982c, 1983). Although the laboratory studies have helped establish causality, they have been criticized on conceptual grounds. Most notably the laboratory studies failed to operationalize the DI and DA approaches as the group structural interventions

intend by their original proponents (Mitroff, 1982a,b; Mitroff and Mason, 1981; Schweiger and Finger, 1984; Schweiger *et al.*, 1986; Chanin and Shapiro, 1985).

Attempting to improve upon operational deficiencies of previous laboratory studies, Schweiger *et al.* (1986) used a sample of MBA students to investigate the comparative effectiveness of DI, DA and a third approach, Consensus (C), as *group* structural interventions. Consensus was included in the study due to its pervasive use in organizations and management training programs (Bourgeois, 1980; *Business Week*, 1978; Mintzberg, Raisinghani and Theoret, 1976; Quinn, 1980; Hall, 1971; Nemiroff, Pasmore and Ford, 1976). Consensus encourages open discussion among group members but does not formally structure or encourage conflict. Schweiger *et al.* (1986) found both DI and DA groups to be superior to consensus group with respect to the quality of recommendations and the validity and importance of assumptions. However, DI groups were found to be superior to DA groups only with respect to the validity of assumptions, and not at all with respect to the quality of recommendations.

An important question raised, but not answered, by Schweiger *et al.* (1986) concerned why performance difference existed among the three approaches. Although the proponents of each approach have theorized about such differences there has been no empirical research to confirm or refute their claims. If we are to understand fully the value of these approaches, then researchers must directly examine how they function and utilize the capabilities of group members. A major premise underlying the use of groups in strategic decision-making is that groups can draw on the knowledge and perspectives of many people. As Mason and Mitroff (1981: 13-14) note:

the raw material for forging solutions to wicked problems is not concentrated in a single head, but rather is widely dispersed among the various parties at stake. For any given wicked problem there is a variety of classes of expertise. Every affected party is an expert on some aspect of the problem and its solution.

Thus, it is logical to expect that the quality of a particular approach is a function of its ability to utilize the capabilities of individual group members (Nemiroff *et al.*, 1976).

There are two methods for assessing group utilization of individual capabilities (Nemiroff *et al.*, 1976). The first measures the difference between a group's performance and the average performance scores of its members; this difference reflects how effectively the group used the capabilities of its members in arriving at recommendations and assumptions. The second measures whether a group's performance exceeds that of its best member. Both measures focus on the synergistic effects of group member interaction. Synergy is said to exist when a group's performance surpasses that of the capabilities of its group members. This occurs when group members effectively exchange, constructively criticize, and build upon each other's ideas (Davis, 1969). Synergy is the major explanation offered for the superiority of DI over the other approaches (Mason and Mitroff, 1981). Using group data analyzed in Schweiger *et al.* (1986) and individual group member data not previously analyzed, this paper will test the following hypotheses:

H1: The performance of DI, DA and C groups will be better than the average performance of individuals in the respective groups.

H2: The differences between group and average individual performance will vary among groups using different approaches.

(a) The difference between group performance and average individual performance will be greater for DI groups than DA or C groups.

(b) The difference between group performance and average individual performance will be greater for DA groups than C groups.

H3: The performance of DI, DA and C groups will be better than that of the best individual in the respective groups.

H4: There will be differences in the extent to which groups using different approaches exceed the performance of the best individual in the group.

(a) DI groups more often will exceed the performance of the best individual in the group than will DA or C groups.

(b) DA groups more often will exceed the performance of the best individual in the group than will C groups.

METHOD

Subject and procedures

Participants in this study were 120 MBA students at a Southwestern university. Their mean age was 29 years ($SD = 5.2$); their mean number of years of full-time work experience was 6.4 ($SD = 4.9$); and their mean number of years of managerial experience was 2.3 ($SD = 3.9$). With respect to managerial experience, 45.8 percent had never been managers, 34.2 percent had been lower-level managers, 15.8 percent had been middle-level managers and 3.3 percent had been top-level managers. Seventy-two were men and 48 were women. Although most of the students were not strategic decision-makers there is reason to believe that their education and managerial experience would lead them to think and act much as would 'real-world' managers.

The study was conducted in three evening sections of a semester-long business policy course. This course was chosen because the decisions required in it were similar to those for which the DI, DA and C approaches were designed. Data collection began at mid-semester to allow time for the students to have developed the knowledge necessary for making strategic decisions. Prior to the start of the research, students have been exposed to 7 weeks of lectures, readings, and case studies focused on strategic management. Participation was a graded course requirement and the study took place during and outside of class time. Identical procedures were used in each section; they are fully described elsewhere (Schweiger *et al.*, 1986).

Task

The Leitch Quality Drug Company case (Glueck, 1980) was used as the decision task in this study. We chose this relatively brief case because it presents a number of strategic problems yet allows analysis in a reasonable amount of time.

Moreover, it allowed us to assess causality while building realism into our study. This is consistent with the recommendations of Schwenk (1982a) and Nees (1983). The case describes a drugstore company that faced significant environmental changes, including demographic shifts and new competition from large super-discount drugstores. Each of its three stores operated in very different neighborhood markets. Internal

problems included weak accounting systems, poor inventory control, and questionable pricing policies.

Each subject submitted written recommendations and the assumptions they depend on prior to the formation of groups. These were used as measures of individual group member performance. Following this individuals were randomly assigned to one of 30 four-person groups. Groups were then randomly assigned to one of three manipulations as described below. Individuals and groups were instructed to analyze the company's situation and address recommendations to one of the owners. They were also instructed to present all the facts and assumptions that supported their recommendations.

Experimental manipulation

Three manipulations, developed in Schweiger *et al.* (1986), were used in this study: DI, DA and C. They are presented below:

Dialectical inquiry manipulation

Ten groups received the following statement of information and instruction:

In the dialectical inquiry approach two opposing recommendations, based on contrary assumptions, are developed from the same data. The two recommendations and their respective assumptions are subjected to an in-depth, critical evaluation through a debate between two advocate subgroups. Using the same data, the debaters attempt to spell out the implications of each decision, reveal its underlying assumptions, and challenge (or defend) those assumptions as effectively as possible. In other words, each side is trying to win the debate. Following the debate the two advocate subgroups should settle on which assumptions survived the scrutiny of debate and attempt to develop a recommendation based on them. Proponents of the dialectical inquiry decision-making approach believe that sound judgements or recommendations are more likely to result from thorough identification and criticism of proposed decisions and their underlying assumptions. Typical guidelines and procedures used in the dialectical inquiry approach follow:

1. Divide your four-person group into two (2) two-person advocate subgroups.
- 2a. One subgroup should develop recommendations and build an argument for them, supported by all key assumptions, facts, and data that underlie them. Write the

- recommendations, assumptions, facts, and data as completely and lucidly as you can on the Subgroup 1 Recommendations form provided for this purpose. Return the form to the other subgroup.
- 2b. The other subgroup should await receipt of the list of key assumptions made by the first group. While waiting, you may discuss the case only between yourselves. Upon receiving this list of key assumptions, the second subgroup should develop plausible assumptions which negate the first subgroup's key assumptions. Using these new assumptions, the second subgroup should develop counter-recommendations and complete the Subgroup 2 Recommendations form.
 3. Following step 2b, both advocate subgroups should present both orally and in writing their assumptions, recommendations, and supporting facts and data to the other subgroup.
 4. The two advocate subgroups debate their recommendations and the validity of the assumptions they have made. The objective of this data is to arrive at a final list of assumptions that is acceptable to both subgroups.
 5. Once the debate is completed, you should reach agreement on which assumptions survived. Any new assumptions that arise from the debate should also be included.
 6. Using the surviving assumptions, develop recommendations.
 7. Record the final recommendations, assumptions, facts, and data on the Final Recommendations form provided.

Devil's advocacy manipulation

Ten groups received the following statement of information and instruction:

The devil's advocacy approach develops a solid argument for a reasonable recommendation, then subjects that recommendation to an in-depth, formal critique. The critique calls into question the assumptions and recommendations presented to the devil's advocate, and attempts to show why the recommendations should not be adopted. Through repeated criticism and revision, the approach leads to mutual acceptance of a recommendation. Proponents of this decision-making approach believe that good recommendations and assumptions will survive even the most forceful and effective criticism and that this approach is more likely to yield sound judgements or recommendations. Here are some guidelines and procedures to follow in using the devil's advocate approach.

1. Divide your four-person group into two (2) two-person subgroups. Assign one

2. Discuss the Leitch Quality Drugs case with your subgroup partner.
- 3a. The subgroup that is not devil's advocate then should develop a set of recommendations and build an argument for them, supported by all key assumptions, facts, and data that underlie them. Write the recommendations, assumptions, facts, and data as clearly and thoroughly as you can on the Subgroup 1 Recommendations form provided for this purpose.
- 3b. Meanwhile, the devil's advocate subgroup should prepare for their critique by discussing the case and any critical assumptions, data, etc., which they can identify.
4. The first subgroup presents its written recommendations and assumptions to the devil's advocate subgroup. The devil's advocate subgroup subjects the recommendations to a formal critique. The critique attempts to uncover all that is wrong with the recommendations, assumptions, facts, and data and to expound the reasons why the recommendations should not be adopted.
5. Following Step 4, the critique is presented to the first subgroup orally and on the Critique form provided. The first subgroup then meets separately once again and revises its recommendations to satisfy the valid criticisms of the devil's advocate subgroup.
6. Repeat Steps 4 and 5 until both subgroups can accept the recommendations, assumptions, and data.
7. Write the final recommendations, assumptions, facts, and data on the Final Recommendations form with which you have been provided.

Consensus manipulation

Ten groups received the following statement of information and instruction:

The consensus approach relies on a thorough, open, and constructive discussion and examination of the recommendations and underlying assumptions developed individually by the group members. In the course of this discussion each group member should have the opportunity to present his recommendations, the underlying assumptions, and relevant facts and data in as clear and logical a manner as possible. Through discussion, questioning, and more complete exchange of information and opinion, the group seeks a better recommendation than might be produced by a single person. It is not necessary that each person be completely satisfied with the assumptions and recommendations — only

that each can accept them on the basis of logic and a willingness to consider them as feasible. *Consensus* is said to exist when *all* group members can accept the assumptions and recommendation on this basis. Proponents of this decision-making approach believe that better assumptions and recommendations result from a more complete investigation and airing of data and ideas and a logical resolution of differences within the group. Here are some guidelines to use in achieving consensus:

1. Avoid arguing blindly for your own assumptions and recommendations. Present your position clearly, logically, and persuasively, but consider carefully the comments and reactions of other group members. If you present the same points again, take comments and reactions into account.
2. Avoid making 'win-lose' statements in your discussion. Discard the notion that someone must win and someone must lose in the discussion. When impasses occur, look for the next most acceptable solution for all parties.
3. Avoid changing your mind simply to avoid conflict and reach agreement. Withstand pressures to yield which have no logically sound foundation. Strive for enlightened flexibility; avoid mere capitulation.
4. Avoid conflict-reducing techniques such as majority voting, tossing a coin, and the like. Differences of opinion indicate an incomplete exchange of relevant information on someone's part; press for additional sharing of task or emotional data where it seems in order.
5. View differences of opinion as natural and helpful rather than as a hindrance to decision-making. Generally, the more assumptions and recommendations expressed, the greater the likelihood of conflict, and the richer the resources used in solving the problem at hand.
6. View all initial agreements as suspect. Explore the reasons for the apparent agreement. Make sure that people have arrived at similar recommendations either for the same reasons or for complementary reasons before incorporating such recommendations into the group's final set.

Dependent measures

Expert judges' ratings of each individual's and each group's recommendations and assumptions were used to assess performance. Judges were aware of the purpose of the experiment and its hypotheses, but had *no* knowledge of the groups

individuals had been assigned to or the conditions (i.e. approaches) groups had been assigned to. It is therefore unlikely that judges' ratings were biased by knowledge of either individual or group assignments. Individual and group performance were scored separately. Two judges, who are educators and researchers in strategic management, independently rated assumptions on three dimensions: (1) the number of assumptions, (2) the validity of the assumptions in light of information provided in the case, and (3) the importance of the assumptions with respect to the decisions made.

Judges computed scores for numbers of assumptions by counting the assumptions that each individual or group provided. They rated the validity of each assumption on a five-point scale ranging from 1 (low confidence in the validity of the assumption) to 5 (high confidence in the validity of the assumption). Validity was defined for judges as the likelihood that the assumption being rated was accurate (given material contained in the case) and would occur (Mason and Mitroff, 1981). Judges rated the importance of each assumption on a five-point scale ranging from 1 (not important) to 5 (very important). Importance was defined for judges as the significance of the assumption on the outcome of the case (Mason and Mitroff, 1981). Final individual and group validity scores were calculated by summing the validity ratings of each individual's or group's assumptions and dividing the sum by the number of assumptions stated by the individual or group. The average individual member score was further computed by summing the performance measure of interest for all members of a group and then dividing it by the number of members. The best individual member score was simply that of the highest individual performance score among a group's members. Judges computed importance scores in the same way. The study used validity and importance averages because otherwise scores of individuals or groups that identified many low-rated assumptions could be inflated beyond those of individuals or groups that identified fewer, but more key assumptions.

The two judges also rated the global quality of each individual's or group's recommendations on a five-point scale. The scale ranged from 1 (low quality) to 5 (high quality). Quality was broadly defined for judges as the extent to which

the recommendations were consistent with the external environment of the firm, and were appropriate in light of the firm's resources; the extent to which parts of the recommendations were internally consistent; and the extent to which the recommendations were workable given the firm's situation (Tilles, 1963).

Manipulation checks

To determine whether the groups used the decision-making approaches, a questionnaire was administered to each member immediately after the group's work was completed. The questionnaire included six items used as manipulation checks. Each item was rated on a five-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). Two items measured the dialectical inquiry manipulation: (1) the two subgroups considered two conflicting sets of recommendations based on opposite sets of assumptions; and (2) on the basis of a debate over two sets of opposing assumptions developed by each subgroup, the group developed a pool of valid assumptions and based its recommendations on these assumptions. Two items measured the devil's advocacy manipulation: (1) one subgroup initially developed a set of recommendations followed by a formal critique of it by a second subgroup; and (2) the group's final set of recommendations was based on the outcomes of a formal critique by one subgroup of the initial recommendations of another subgroup. Two items measured the consensus manipulation: (1) during the group meeting only, members of the group individually presented their personal recommendations and assumptions; and (2) the group's final set of recommendations was the combined, but not necessarily the averaged, outcome of members' individual contributions. Each group's work session was also tape-recorded. Using as criteria the written instructions provided to the groups, two judges listened to the tapes and independently classified each group as DI, DA, or C.

RESULTS

Manipulation checks

Internal consistency estimates using Cronbach's alpha were computed for each of the two-item manipulation checks. The DI, DA and C scales demonstrated reasonable internal consistency

(respectively 0.88, 0.84, and 0.63). Therefore all three checks were retained for further analysis. Agreement with the DI manipulation check differed significantly by condition ($F_{2,117} = 66.7$, $p < 0.001$), with the DI condition differing significantly from those in the other two conditions ($p < 0.01$, planned *t*-test). Agreement with the DA manipulation check also differed significantly by condition ($F_{2,117} = 86.1$, $p < 0.001$), with the DA condition differing significantly ($p < 0.001$) from those in the other two conditions. Finally, agreement with the consensus manipulation check also differed significantly by condition ($F_{2,117} = 66.4$, $p < 0.001$), with the consensus condition differing significantly from those in the other two conditions ($p < 0.001$).

The judges' classifications of the tapes were less supportive of the success of the manipulations. The two judges agreed on only 70 percent (21 of 30) of the classifications. (Of the nine classifications on which they did not agree, seven involved DI or DA groups.) The two judges discussed and resolved all nine disagreements. Then their final classifications were compared to each group's assigned approaches. There was a match on 23 of 30 classifications, or 76 percent. The seven that did not match involved either DI or DA groups. All C groups were accurately classified. Overall, the individual level manipulation checks were very supportive of the success of the manipulations, whereas the taped manipulation checks were only moderately supportive of the manipulations. This discrepancy may be partially due to the priming effects of the manipulation instructions on individuals' questionnaire responses.

Reliability of performance measures

Before using the performance measure, interrater reliability between the independent ratings of the two judges was computed.¹ The results for each measure are discussed below.

Number of assumptions

For number of assumptions, the correlation between the judges was 0.74; they agreed on 76

¹ Subsequent to the conduct of this study the judges' ratings were compared to those of a Senior Vice-President of Corporate Planning and Development and a Director of Strategic Planning for a division of a Fortune 500 company. Their ratings were similar to those presented below for the two academic judges.

percent of the groups. In only one case was there large disagreement between the judges. Thus, it was felt that each judge could serve as an independent rater.

Quality of assumptions

Two measures were used to assess the quality of the 297 unique assumptions that groups or individuals identified: validity and importance. Interrater correlations and percentage of agreement between judges were computed for both measures of quality. For the validity measure the correlation was 0.88, with full agreement on 66 percent of the judgements. The correlation for the importance measure was 0.78, with full agreement on 59 percent of the judgements. Given the difficulty of rating so many assumptions, we concluded that a reasonable level of interrater reliability had been established to retain both measures. Because the two judges' ratings had only small differences (Schweiger *et al.*, 1986), we averaged them to arrive at final validity and importance ratings of each assumption.

Quality of recommendations

We computed both correlation and percentage of agreement between the two judges' ratings of each group's recommendations. The correlation between the judges was 0.68, with full agreement on 60 percent of the judgements. Because the judges differed substantially on only six groups we used discussion rather than averaging to resolve their differences (Schweiger *et al.*, 1986). All discrepancies were resolved. Based on the high degree of agreement between the judges on all the judgements, one of the judges subsequently rated the recommendations submitted individually by each of the 120 subjects.

Performance²

Table 1 presents means and standard deviations for group and average group member performance. Tables 2, 3 and 4 summarize the results of

² Because the results of the taped manipulation checks were questionable, we conducted analyses parallel to the ones presented in this section without the seven groups for which the judges' classifications did not agree with the actual manipulations. All analyses excluding the seven groups were identical to those of the full sample of 30 groups, except for the validity of assumptions difference scores between the DI and DA groups. It was not significant in the smaller sample.

the performance comparisons discussed in the following sections.

Group vs. average member performance

With respect to the number of assumptions identified, the only significant difference was for the DI groups vs. their average member performance ($F_{1,18} = 17.72, p < 0.01$). An inspection of the means in Table 1 indicates that, in this case, the average member was higher than the group.

The correlation between the validity and importance of assumptions measures was 0.77 ($p < 0.001$) for groups and 0.56 ($p < 0.001$) for individuals. These correlations and conceptual evidence suggesting that these measures are related required that we analyze them together. A one-way MANOVA yielded significant differences between group and average member performance for the DI ($F_{2,117} = 19.02$, Wilk's lambda = 0.309, $p < 0.001$) and DA groups ($F_{2,17} = 4.21$, Wilk's lambda = 0.669, $p < 0.05$). Consensus groups did not do significantly better than their average member performance. To further understand the univariate effects we computed one-way ANOVAs for each dependent variable. For the DI groups both validity ($F_{1,18} = 7.01, p < 0.05$) and importance ($F_{1,18} = 8.15, p < 0.05$) were also significantly better.

The quality of recommendations for the DI groups ($F_{1,18} = 17.72, p < 0.01$) and DA groups ($F_{1,18} = 6.92, p < 0.05$) were significantly better than their average member. Again, the C groups did not do better than their average member.

The results of these analyses provide partial support for Hypothesis 1. Although the DI and DA groups did better than the average of their members, the C groups did not. It is important to note, however, that in no case did a C group do *worse* than its members' average. The point is not that any of the three approaches negated its members' capabilities, but that the DI and DA used them better than did the C approach.

To test Hypothesis 2 we further analyzed the abilities of the DI, DA, and C groups to utilize the capabilities of their members. We computed difference scores between each group and the average of its respective members and then compared these scores across group conditions. A one-way MANOVA indicated significant differences among the conditions in the combined validity and importance of assumptions surfaced

Table 1. Means and standard deviations for group and individual performance

| Independent variables | Dependent variables | | | | | | | |
|-----------------------|-----------------------|------|-------------------------|------|---------------------------|------|----------------------------|------|
| | Number of assumptions | | Validity of assumptions | | Importance of assumptions | | Quality of recommendations | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Dialectical inquiry | | | | | | | | |
| Group | 12.50 | 3.57 | 4.23 | 0.38 | 4.04 | 0.29 | 4.05 | 0.28 |
| Average of members | 16.27 | 1.89 | 3.35 | 0.32 | 3.22 | 0.46 | 3.51 | 0.29 |
| Devil's advocacy | | | | | | | | |
| Group | 12.65 | 6.96 | 3.82 | 0.52 | 3.72 | 0.50 | 3.90 | 0.46 |
| Average of members | 13.57 | 2.81 | 3.35 | 0.26 | 3.20 | 0.09 | 3.46 | 0.24 |
| Consensus | | | | | | | | |
| Group | 13.40 | 5.47 | 3.15 | 0.31 | 3.09 | 0.21 | 3.45 | 0.60 |
| Average of members | 13.14 | 1.93 | 3.25 | 0.28 | 3.25 | 0.23 | 3.16 | 0.20 |

Table 2. Results of performance comparisons between groups and average of individual group members

| Dependent measures | Results |
|-----------------------------------|--------------------------|
| <i>Quality of assumptions</i> | |
| MANOVA | |
| Validity and importance | |
| DI vs. group average | $F_{2,17} = 19.02^{***}$ |
| DA vs. group average | $F_{2,17} = 4.21^*$ |
| C vs. group average | $F_{2,17} = 1.66$ |
| ANOVA | |
| Validity | |
| DI vs. group average | $F_{1,18} = 30.93^{***}$ |
| DA vs. group average | $F_{1,18} = 7.01^*$ |
| C vs. group average | NA |
| Importance | |
| DI vs. group average | $F_{1,18} = 21.86^{***}$ |
| DA vs. group average | $F_{1,18} = 8.15^*$ |
| C vs. group average | NA |
| Number of assumptions | |
| DI vs. group average | $F_{1,18} = 8.70^{**}$ |
| DA vs. group average | $F_{1,18} = 0.15$ |
| C vs. group average | $F_{1,18} = 0.02$ |
| <i>Quality of recommendations</i> | |
| DI vs. group average | $F_{1,18} = 17.72^{**}$ |
| DA vs. group average | $F_{1,18} = 6.92^*$ |
| C vs. group average | $F_{1,18} = 2.07$ |

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; NA = Not applicable.

($F_{4,52} = 5.76$, Wilk's lambda = 0.480, $p < 0.01$). Using one-way ANOVAs to examine the univariate effects we found significant differences for both validity ($F_{2,27} = 10.18$, $p < 0.01$) and importance ($F_{2,27} = 8.36$, $p < 0.01$). As indicated

in Table 3, all of the planned t -tests among the groups were significant except for the comparison between the DI and DA groups for importance of assumptions.

A one-way ANOVA for number of assumptions

Table 3. Results of difference scores between groups and average individual group members across group conditions

| Dependent measure | Results |
|-----------------------------------|-------------------------|
| <i>Quality of assumptions</i> | |
| MANOVA | |
| Validity and importance | $F_{4,52} = 5.76^{**}$ |
| ANOVA | |
| Validity | $F_{2,27} = 10.18^{**}$ |
| Importance | $F_{2,27} = 8.36^{**}$ |
| Number of assumptions | $F_{2,27} = 2.36$ |
| <i>t-tests</i> | |
| Validity | |
| DI vs. DA | $t_{18} = 3.29^{**}$ |
| DI vs. C | $t_{18} = 4.31^{***}$ |
| DA vs. C | $t_{18} = 2.84^{**}$ |
| Importance | |
| DI vs. DA | $t_{18} = 0.92$ |
| DI vs. C | $t_{18} = 3.69^{**}$ |
| DA vs. C | $t_{18} = 3.71^{**}$ |
| <i>Quality of recommendations</i> | |
| ANOVA | |
| Quality of recommendations | $F_{2,27} = 5.96^*$ |
| <i>t-tests</i> | |
| Quality of recommendations | |
| DI vs. DA | $t_{18} = 0.28$ |
| DI vs. C | $t_{18} = 3.52^*$ |
| DA vs. C | $t_{18} = 2.90^*$ |

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 4. Group performance vs. best group member performance

| | Number of assumptions | Validity of assumptions | Importance of assumptions | Quality of recommendations |
|--|-----------------------|-------------------------|---------------------------|----------------------------|
| Proportion of groups surpassing best member performance | | | | |
| DI | 0/10 = 0% | 8/10 = 80%* | 8/10 = 80%* | 8/10 = 80%* |
| DA | 3/10 = 30% | 5/10 = 50% | 6/10 = 60% | 8/10 = 80%* |
| C | 2/10 = 20% | 2/10 = 20% | 1/10 = 10% | 2/10 = 20% |
| Results of proportional comparisons among group conditions | | | | |
| DI vs. DA | NS | NS | NS | NS |
| DI vs. C | NS | $p < 0.05$ | $p < 0.05$ | $p < 0.05$ |
| DA vs. C | NS | NS | $p < 0.05$ | $p < 0.05$ |

* $p < 0.05$; NS = not significant.

yielded no significant differences. With respect to quality of recommendations, however, significant differences were found across conditions ($F_{2,27} = 5.96$, $p < 0.05$). Further analyses using *t*-tests (see Table 3) found significant differences between the DI and C groups and between the

DA and C groups, but not between the DI and DA groups. These findings fully support Hypothesis 2b but only partially support Hypotheses 2 and 2a. Essentially the DI and DA groups utilized average member capabilities significantly better than did the C groups.

To determine whether the DI, DA, and C groups significantly exceeded the performance of their best group member, the performance score for each group was compared to that of its respective best member. A proportion of the number of groups that exceeded the best member to the total number of groups for each condition was calculated. Tests of significance of these proportions were then performed. These analyses used 0.33 as the expected value of a group's ability to exceed the performance of its best member, based on data presented by Nemiroff *et al.* (1976). Analyses of the data reported in Table 4 indicated that the DI groups significantly exceeded their best member for validity of assumptions ($p < 0.05$), importance of assumptions ($p < 0.05$), and quality of recommendations ($p < 0.05$); whereas the DA groups exceeded their best members only for quality of recommendations ($p < 0.05$). No other findings were significant; thus only partial support was provided for Hypothesis 3.

To determine whether the preceding proportions were significantly different among group conditions (i.e. did groups in one condition more often exceed the performance of their best member than in another condition), tests of significance for differences between two proportions were performed (Bruning and Kintz, 1968). The results of these analyses, presented in Table 4, indicate that for importance of assumptions ($p < 0.05$) and quality of recommendations ($p < 0.05$) the proportions for the DI and DA conditions were significantly greater than the proportions for the C condition. For validity of assumptions only the proportion for the DI condition was significantly greater than the proportion for the C condition ($p < 0.05$). No differences were found between the DI and DA groups; thus only partial support was found for Hypotheses 4, 4a, and 4b.

DISCUSSION

Dialectical inquiry and devil's advocacy have received considerable attention in the literature concerning their ability to improve the quality of decisions made by top management groups. Although numerous studies have attempted to

examine the comparative effectiveness of these two approaches, they have been criticized on conceptual and methodological grounds. Most notably, the studies have failed to operationalize DI and DA as group approaches, for which their proponents originally intended.

Attempting to improve upon previous research, Schweiger *et al.* (1986) conducted the first controlled laboratory experiment to assess the comparative effectiveness of DI, DA, and a third approach C, as group approaches. The study found consistent significant differences in effectiveness among the approaches, but provided no data to explain why. The present study attempted to extend the Schweiger *et al.* (1986) study by examining the extent to which the approaches utilized the capabilities of individual group members in forging group strategic decisions. Such an examination allowed us to evaluate the major premises underlying the value of group versus individual decision-making and the superiority of particular group approaches. We believe that this is an important first step in understanding how and why these approaches work.

The results of this study provide evidence to support these major premises. In particular, they indicate that the DI and DA groups exceeded the average performance of their respective members, whereas the C groups did not. Moreover, this finding was confirmed by the comparative differences between group and average member performance among the three approaches. The differences achieved by both DI and DA groups were significantly greater than those achieved by C groups for most performance measures. For one performance measure, validity of assumptions, DI groups had greater differences than DA groups. Of additional interest is that the group average for number of assumptions was significantly greater than that of the group for DI. This suggests that DI groups generate fewer, but better, assumptions. It seems likely that the debate process inherent in the DI approach is very effective at evaluating and paring down during group deliberations the assumptions developed by individuals.

Comparing groups versus average member performance, however, provides only partial insight into how well an approach utilizes members' capabilities. The true test of a group approach is whether its assumptions and recommendations are superior to those of the

best member. Our results with respect to the performance of groups versus that of their best member is quite revealing. Only the DI approach demonstrated consistent group superiority across all performance measures. The DA approach showed consistent superiority only with respect to quality of recommendations, although DA groups exceeded their best member at least half the time with respect to the validity and importance of assumptions. The results of the C groups, on the other hand, were disappointing. Comparisons across the approaches further revealed that both DI and DA groups exceeded their best member significantly more often than the C approach.

The results did not reveal consistent differences in member utilization between the conflict-inducing approaches. Therefore, we tentatively conclude that the evidence does not support the claims of Mason and Mitroff (1981) and Mason (1969), who have argued that DI is superior to DA. The results do partially support the findings of Cosier and his colleagues (Cosier, 1981, 1982, 1983; Schwenk, 1982a,c, 1983) that the DI approach is not necessarily superior to the DA approach. Moreover, the results are also consistent with their findings that approaches that formally structure conflict are superior with respect to decision quality to those approaches that do not.

Our study provides only a partial understanding of the workings of the DI and DA approaches. We have established that these two approaches do not substantially differ in their utilization of members' capabilities. However, additional research is needed to determine whether and how certain aspects of the debate process (e.g. critical questioning, organizing presentation of conflicting ideas) affect strategic decision-making. Such research ought to examine not only each aspect's impact on decision quality, but also its impact on participants' affective responses to the group, the process, and the decision. Previous research has shown that DI and DA approaches, when compared to the C approach, leave members less committed to group decisions and less willing to work again with their group (Schweigèr *et al.*, 1986). These may affect the ability of an ongoing group to continue to function effectively over time.

It would be useful to examine each aspect of the group process, assessing its contributions to

decision quality *and* to undesirable effects on participants. The debate might be restructured to eliminate aspects that are both nonproductive and dysfunctional. Also, participants might be trained so as to reduce the dysfunctional effects of aspects that *are* productive.

As does any study, this one has its limitations. The major limitations concern the generalizability of the findings to actual strategic decision-making situations and thus warrant caution in interpreting the 'real-world' implications of the present findings. Although a strategic decision-making case was the task, it had certain deficiencies. Since the company in the case was privately held, subjects received the information to analyze the situation from one source. In most strategic decision situations, however, decision-makers have to learn what information is needed and where to find it. In addition, those actually involved in running a company would already have had certain industry, company, and task group experiences that subjects in this study did not have.

A second limitation was the subjects used. Most subjects were not real-life strategic decision-makers, although the demographic data suggested that they had a reasonable degree of managerial experience.

Moreover, the results of the study also provide some inductive evidence that the sample was reasonable. For example, the means for DI for validity and importance of assumptions and for quality of recommendations all exceeded 4 on a five-point scale. Both the overall average of the individuals and the average for the consensus approach (the weakest of the three approaches) exceeded 3.0 for all dependent measures. These results suggest that the responses of individuals and groups were rated toward the high-quality end of the continuum by the judges. The fact that the ratings were consistent with those of a Senior Vice-President of Corporate Planning and Development and a Director of Strategic Planning for a *Fortune* 500 company (see footnote 2) suggests that subjects made high-quality, non-naive assumptions and recommendations.

Finally, any naiveté on the part of the students would have a *conservative* effect on the results. To the extent that subjects could not fathom the strategic issues in the case, or follow the DI or DA instructions for arriving at group recommendations, differences among the approaches would

tend to be understated rather than overstated. The results indicate that the DI and DA approaches made better use of the best assumptions and recommendations brought to the group by individual members. It might be argued that the subjects possessed less strategic expertise than 'real-world' managers, yet if true this would reduce (not enhance) the likelihood of finding these differences.

A third limitation is the duration of the study. The experiment was conducted over a 1-week period. Although top managers may make some decisions in such a short period, it is likely that comprehensive strategic decisions, such as were required here, would have required significantly more time. Thus the duration of the study may have affected the results.

Fourth, the study did not include time as a performance variable. Groups operated within a fixed period and had little incentive to finish more quickly. In practice, however, top management's time carries a high opportunity cost. Moreover, speed may be crucial in some decisions. The best decision-making approach, therefore, may depend on the expected value of incremental decision quality, the cost of additional management work hours devoted to the decision, and the urgency of the situation.

Finally, this study, like most group decision-making studies conducted in laboratory settings, used ad hoc groups divorced from an organizational context. Clearly, longitudinal studies of actual strategic decision-making groups are needed. Such studies would provide additional insight into the long-term effects of programmed group conflict on such groups. In particular, recurrent conflict may eventually undermine the ability of a group to work together, resulting in poor utilization of members' capabilities. A second contribution of longitudinal studies could be to identify the effects of learning on individual capabilities and any additional synergy brought about by increased experience with these group decision approaches. Learning and experience could influence the relative effectiveness of these approaches over time. A third contribution of such studies would be to examine the effects that group norms and political processes, and the development of shared information, perceptions, and values among group members, might have on the adoption and sustained use of group conflict approaches (Beyer, 1981). These studies

should help us understand and organizational conditions that affect the utilization and effectiveness of the group approaches.

REFERENCES

- Beyer, J. M. 'Ideologies, values, and decision making in organizations'. In Nystrom, P. and W. H. Starbuck, (eds.), *Handbook of Organizational Design*, Vol. 2, Oxford University Press, London, 1981, pp. 166-197.
- Bourgeois, L. J. 'Performance and consensus', *Strategic Management Journal*, 1, 1980, pp. 227-248.
- Bruning, J. L. and A. L. Kintz. *Computational Handbook of Statistics*, Scott Foresman, Glenview, IL, 1968.
- Business Week*. 'Olin's shift to strategy planning', 27 March 1978, pp. 102-105.
- Chanin, M. N. and H. J. Shapiro. 'Dialectical inquiry in strategic planning: Extending the boundaries', *Academy of Management Review*, 10, 1985, pp. 663-675.
- Cosier, R. A. 'The effects of three potential aids for making strategic decisions on prediction accuracy', *Organizational Behavior and Human Performance*, 22, 1978, pp. 295-306.
- Cosier, R. A. 'Inquiry method, goal difficulty, and context effects on performance', *Decision Sciences*, 11, 1980, pp. 1-16.
- Cosier, R. A. 'Dialectical inquiry in strategic planning: a case of premature acceptance'? *Academy of Management Review*, 6, 1981, pp. 643-648.
- Cosier, R. A. 'Methods for improving the strategic decision: dialectic versus the devil's advocate', *Strategic Management Journal*, 3, 1982, pp. 373-374.
- Cosier, R. A. 'Approaches for experimental examination of the dialectic', *Strategic Management Journal*, 4, 1983, pp. 79-84.
- Cosier, R. A. and J. C. Aplin. 'A critical view of dialectical inquiry as a tool in strategic planning', *Strategic Management Journal*, 1, 1980, pp. 343-356.
- Cosier, R. A. and P. L. Rechner. 'Inquiry method effects on performance in a simulated business environment', *Organizational Behavior and Human Decision Processes*, 36, 1985, pp. 79-95.
- Cosier, R. A., T. A. Ruble and J. C. Aplin. 'An evaluation of the effectiveness of dialectical inquiry systems', *Management Science*, 24, 1978, pp. 1483-1490.
- Davis, J. H. *Group Performance*. Addison-Wesley, Reading, MA, 1969.
- Glueck, W. F. *Business Policy and Strategic Management*, 3rd edn. McGraw-Hill, New York, 1980.
- Hall, J. 'Decisions, decisions, decisions', *Psychology Today*, 5, 1971, pp. 51-54, 86, 88.
- Lourenco, S. V. and J. C. Glidewell. 'A dialectical approach analysis of organizational conflict', *Administrative Science Quarterly*, 20, 1974, pp. 489-508.
- Mason, R. O. 'A dialectical approach to strategic planning', *Management Science*, 15, 1969, pp. B403-B414.
- Mason, R. O. and I. I. Mitroff. *Challenging Strategic*

- Planning Assumptions**, Wiley-Interscience, New York, 1981.
- Mintzberg, H., D. Raisinghani and A. Theoret. 'The structure of unstructured decision processes', *Administrative Science Quarterly*, **21**, 1976, pp. 246-275.
- Mitroff, I. I. 'Talking past one's colleagues in matters of policy', *Strategic Management Journal*, **3**, 1982a, pp. 374-375.
- Mitroff, I. I. 'Dialectic squared: a fundamental difference in perception of the meanings of some key concepts in social science', *Decision Sciences*, **13**, 1982b, p. 222-224.
- Mitroff, I. I. and J. R. Emshoff. 'On strategic assumption-making: a dialectical approach to policy and planning', *Academy of Management Review*, **4**, 1979, pp. 1-12.
- Mitroff, I. I. and R. O. Mason. 'The metaphysics of policy and planning: a reply to Cosier', *Academy of Management Review*, **6**, 1981, pp. 649-651.
- Mitroff, I. I., V. P. Barabba and R. H. Kilmann. 'The application of behavioral and philosophical technologies to strategic planning: a case study of a large federal agency', *Management Science*, **24**, 1977, pp. 44-58.
- Nees, D. B. 'Simulation: a complementary method for research on strategic decision making processes', *Strategic Management Journal*, **4**, 1983, pp. 175-185.
- Nemiroff, P. M., W. A. Pasmore and D. L. Ford. 'The effects of two normative structural interventions on established and ad hoc groups: implications for improving decision making effectiveness', *Decision Sciences*, **7**, 1976, pp. 841-855.
- Quinn, J. B. *Strategies for Change: Logical Incrementalism*, Richard B. Irwin, Homewood, IL, 1980.
- Schweiger, D. M. and P. A. Finger. 'The comparative effectiveness of dialectical inquiry and devil's advocacy: the impact of task biases on previous research findings', *Strategic Management Journal*, **5**, 1984, pp. 335-350.
- Schweiger, D. M., W. R. Sandberg and J. W. Ragan. 'Group approaches for improving strategic decision making: a comparative analysis of dialectical inquiry, devil's advocacy, and consensus', *Academy of Management Journal*, **29**, 1986, pp. 51-71.
- Schwenk, C. R. 'Why sacrifice rigor for relevance? a proposal for combining laboratory and field research in strategic management', *Strategic Management Journal*, **3**, 1982a, pp. 213-225.
- Schwenk, C. R. 'Effects of inquiry methods and ambiguity tolerance on prediction performance', *Decision Sciences*, **13**, 1982b, pp. 207-221.
- Schwenk, C. R. 'Dialectical inquiry in strategic decision making: A comment on the continuing debate', *Strategic Management Journal*, **3**, 1982c, pp. 371-373.
- Schwenk, C. R. 'Laboratory research on ill-structured decision aids: The case of dialectical inquiry', *Decision Sciences*, **14**, 1983, pp. 140-144.
- Schwenk, C. R. 'Effects of planning aids and presentation media on performance and affective responses in strategic decision making', *Management Science*, **30**, 1984, pp. 263-272.
- Schwenk, C. R. and R. A. Cosier. 'Effects of the expert, devil's advocate, and dialectical inquiry methods on prediction performance', *Organizational Behavior and Human Performance*, **26**, 1980, pp. 409-424.
- Tilles, S. 'How to evaluate corporate strategy', *Harvard Business Review*, **41**, 1963, pp. 111-121.